15: Hypothesis Tests

Stat 120 | Fall 2025

Prof Amanda Luby

```
library(tidyverse)
library(broom)
library(patchwork)
library(CarletonStats)
```

Do heart attack victims have higher cholesterol levels than non-heart attack victims? The dataset `Cholesterol.csv` has cholesterol measurements (mg/dL) for a sample of heart attack victims (4 days after the heart attack) and a sample of non-heart attack victims.

```
Cholesterol <- read.csv("http://math.carleton.edu/Stat120/RLabManual/Cholesterol.csv")
```

0. Load the data and check the "spreadsheet view". What is each case? How many variables are there?

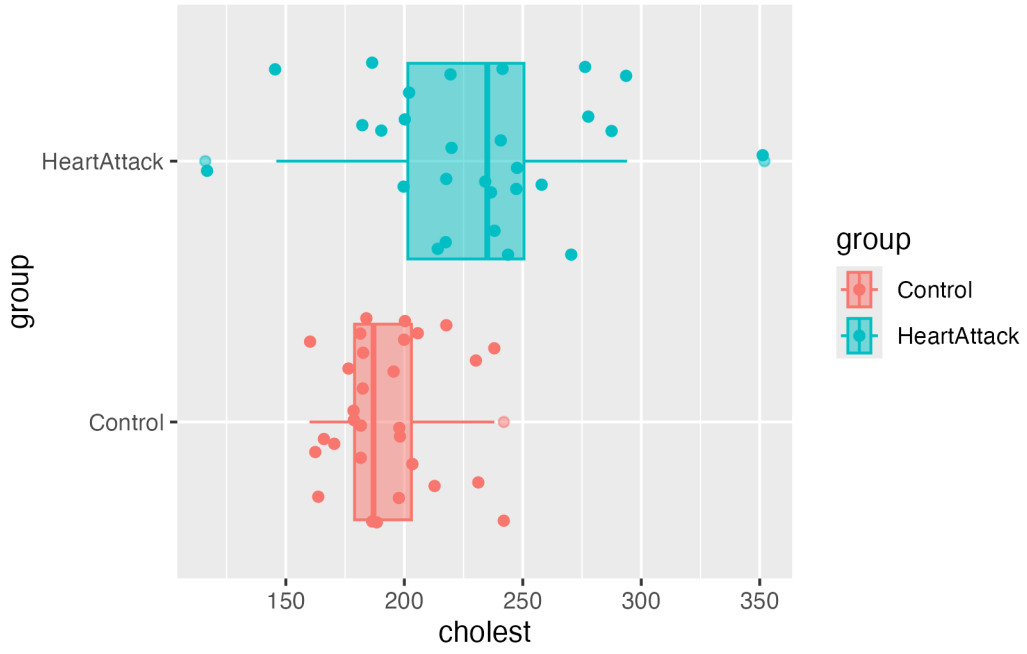
Solution

Each case is a patient. There are 2 variables (group + cholesterol, plus a column for patient ID)

1. Create an appropriate graph to answer the research question.

Solution

```
ggplot(Cholesterol, aes(y = group, x = cholest, fill = group, col = group)) +
  geom_boxplot(alpha = .5) +
  geom_jitter()
```



2. *Without performing a formal statistical test* Does there appear to be evidence of a difference in cholesterol levels between the control group and the heart attack group?

Solution

Yes, it looks like there is a difference in cholesterol between the two groups, although the sample size is kind of small

3. Write out appropriate null and alternative hypotheses using the original research question.

Solution

$$H_0 : \mu_{\text{HeartAttack}} - \mu_{\text{Control}} = 0$$

$$H_A : \mu_{\text{HeartAttack}} - \mu_{\text{Control}} > 0$$

4. Use the `permTest()` command in R to test your hypotheses. (Make sure to include `seed`!)

Solution

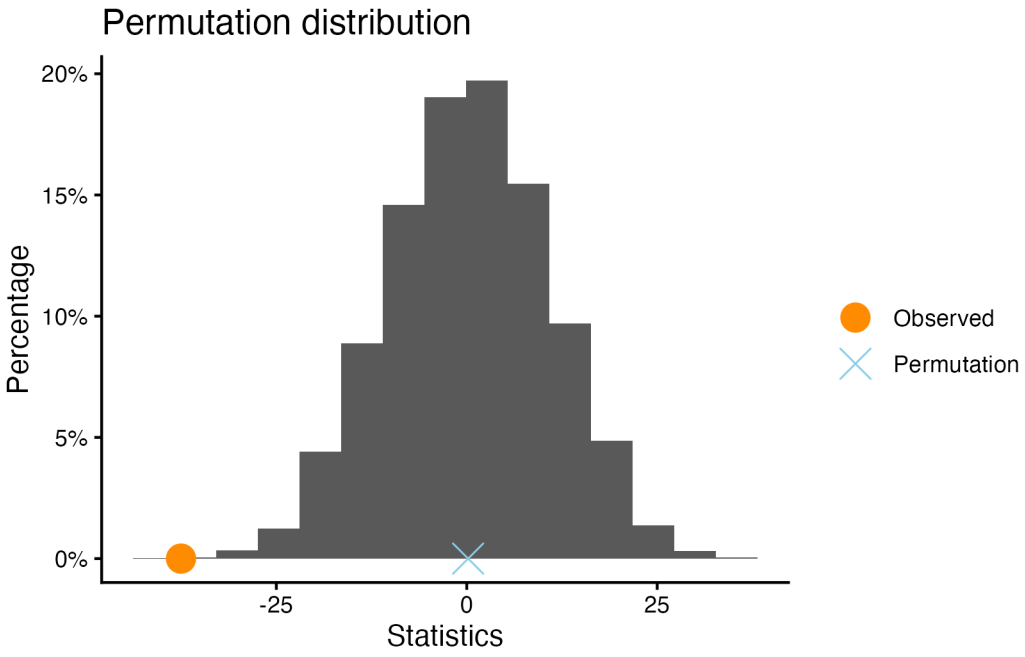
```
permTest(cholest ~ group, data = Cholesterol, alternative = "less", seed = 10156)
```

```
    ** Permutation test **
```

```
Permutation test with alternative: less
Observed statistic
 Control :  193.1333    HeartAttack :  230.6429
Observed difference: -37.50952

Mean of permutation distribution: 0.17994
Standard error of permutation distribution: 10.57276
P-value:  3e-04

   *-------------*
```



5. Report the *statistic* and *p-value* from the R output.

Solution

Statistic is $\bar{X}_1 - \bar{X}_2 = 37.51$
P value is .00003

6. Make a *formal statistical decision* and *report your conclusion in context.*

Solution

At the $\alpha = .05$ level, we reject H_0 and conclude that cholesterol for heart attack victims is likely higher than non heart attack victims.

7. What type of error could you have made in (6)? Do you know the probability that an error occurred?

Solution

Type I error. α

When you're done: Let Amanda know.

Note: This is based on Lab Manual Ch5 #7

8. True or False? If false, explain why or correct the statement.

(a) If a p-value is 10%, there is a 1 in 10 chance the null hypothesis is correct
(b) When a p-value is extremely small, the result is extremely important
(c) A small p-value means that the result could not possibly have been due to chance
(d) A big p-value means that you do not have strong evidence against the null hypothesis
(e) A p-value is the probability of getting a result more in favor of the alternative hypothesis than the result that you observed, assuming the null hypothesis is true
(f) A p-value is the probability of getting a result more in favor of the alternative hypothesis than the result that you observed, assuming the alternative hypothesis is true

Solution

(a) false
(b) false
(c) false
(d) true
(e) true
(f) false